Aesthetic appreciation of the cultural landscape through social media

Tieskens, Koen F.; Van Zanten, Boris T.; Schulp, Catharina J.E.; Verburg, Peter H.

published in
Landscape and Urban Planning
2018

DOI (link to publisher)
10.1016/j.landurbplan.2018.05.002

document version
Publisher's PDF, also known as Version of record

document license
Article 25fa Dutch Copyright Act

Landscape and Urban Planning 177 (2018) 128–137

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

Research Paper

Aesthetic appreciation of the cultural landscape through social media: An analysis of revealed preference in the Dutch river landscape

Koen F. Tieskens[a,*], Boris T. Van Zanten[a], Catharina J.E. Schulp[a], Peter H. Verburg[a,b]

[a] Institute for Environmental Studies (IVM), Faculty of Earth and Life Sciences, VU University Amsterdam, De Boelelaan 1087, 1081HV Amsterdam, The Netherlands
[b] Swiss Federal Research Institute WSL, Research Unit Landscape Dynamics, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland

ARTICLE INFO

Keywords:
Tourism
Cultural ecosystem services
Recreation
Aesthetic enjoyment
Crowd sourcing
Peri-urban area

ABSTRACT

Aesthetic enjoyment and perception are increasingly recognized as important values of cultural landscapes. The study of these values transcends mere physical attributes of the landscape and requires assessment of its social meaning. In recent years the usage of social media has gained momentum to study the aesthetic preferences and perception of the environment. However, until now the different approaches have not yet been sufficiently combined to provide more in depth understanding of what attracts people in the landscape We propose a robust methodology using social media photos from Flickr and Panoramio to estimate the correlation between landscape attributes and landscape preferences. We combine formal modeling of spatial photo distribution based on the occurrence of landscape elements with content analysis of the photos to pinpoint what it is in a landscape that attracts people. We use the Kromme Rijn Area –a peri-urban area in the center of the Netherlands and a popular recreation area- as case study area. The analysis shows that this area is appreciated by its visitors and residents for the presence of monumental buildings, small water bodies and opportunities for hikes along grasslands. The method successfully linked the structural elements of the landscape with the revealed preferences, providing a way of quantifying the appreciation of the landscape. Qualitative surveys remain essential to study motivations for outdoor recreation, but social media data can be incorporated as evidence of what elements of the landscape are valued, where people are interacting with the landscape, and how these interactions characterize a landscape.

1. Introduction

Cultural landscapes are, besides their role in food production, increasingly recognized and valued as objects of aesthetic beauty (Buijs, Pedroli, & Luginbühl, 2006). Their importance for economic welfare and well-being, through for instance recreation, or sense of place, inspired ample scholarly work on the link between these non-material benefits and the physical landscape attributes (Van Zanten, Verburg, Koetse, & van Beukering, 2014). However, the study of these values transcends mere physical attributes of the landscape and requires assessment of its social meaning within a given context (Lothian, 1999; Plieninger et al., 2015). Traditional stated preference approaches often rely on choice experiments representing different landscape attributes with context specific (manipulated) photographs to gain insight into landscape preferences (e.g. Barroso, Pinto-Correia, Ramos, Surová, & Menezes, 2012; van Berkel and Verburg, 2014; van Zanten, Verburg, Scholte, and Tieskens, 2016b). The on-site employment of photographs in choice modeling is generally regarded as an adequate method to unravel landscape preferences as it allows for keeping external factors such as light and weather conditions equal while manipulating landscape elements present in photos (Steen Jacobsen, 2007). However, photographs are unable to capture the experience people have in a landscape as the photo is imposed by the researcher (Scott & Canter, 1997). Consequently, choice experiments cannot avert suffering from a hypothetical bias (Hanley, Mourato, & Wright, 2001).

The rise of social media has opened up new paths in landscape preference studies. Platforms such as Flickr, Panoramio, and Instagram allow their users to upload photos of the environment and place them on a digital map (Casalegno, Inger, Desilvey, & Gaston, 2013; Wood, Guerry, Silver, & Lacayo, 2013). Together, they provide a publicly available database of volunteered geographic information (VGI) with millions of geo-tagged photos spread all over the world (Goodchild, 2007). One of the main advantages of VGI is that it gives an insight into spatial choices and preferences of people without bias of experiments or surveys (Schlieder & Matyas, 2009). Within the last decade applications of VGI have been numerous and include semantic spatial analysis to study collective understandings of spatial concepts (Hollenstein & Purves, 2010) or using sentiment analysis of Twitter data to analyze

* Corresponding author.
E-mail addresses: Koen.tieskens@vu.nl (K.F. Tieskens), Boris.zanten@vu.nl (B.T. Van Zanten), Nynke.schulp@vu.nl (C.J.E. Schulp), Peter.verburg@vu.nl (P.H. Verburg).

https://doi.org/10.1016/j.landurbplan.2018.05.002
Received 15 June 2017; Received in revised form 23 April 2018; Accepted 1 May 2018
Available online 16 May 2018
0169-2046/

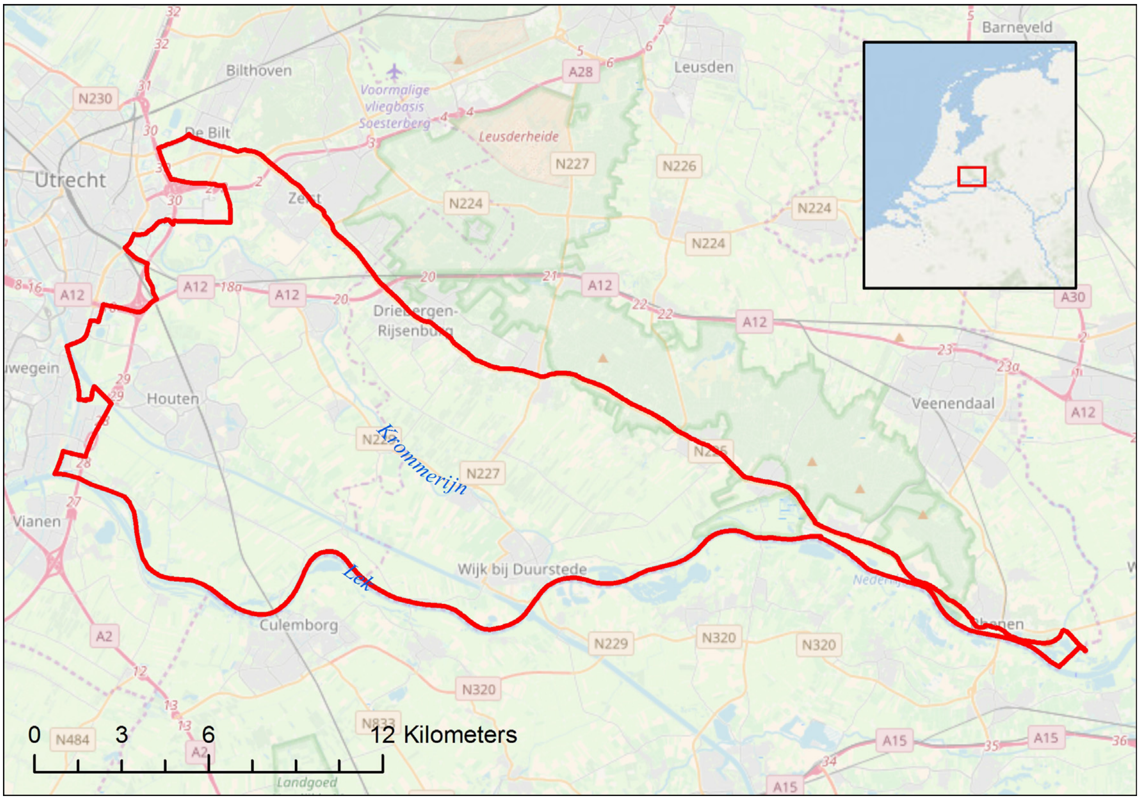

Fig. 1. Location of case study Kromme Rijn area.

presidential election (Gordon, 2013). Recent studies show how VGI allows employing photos generated by users, as proxies for their landscape preferences, rather than hypothetically stated preferences for different landscapes (Gliozzo, Pettorelli, & Haklay, 2016; Hausmann et al., 2017; Oteros-Rozas, Martín-López, Fagerholm, Bieling, & Plieninger, 2017; van Zanten et al., 2016a).

Wood et al. (2013) were among the first to utilize social media content for landscape value research. They found evidence that actual visitation rates can successfully be predicted using the density of geotagged Flickr photos. Where Wood et al. (2013) used the trail of geotagged photos to explain spatial behavior of people, Casalegno et al. (2013) applied a similar method but instead mapped preferences. They used Panoramio densities as a proxy for the aesthetic value of landscapes in Cornwall, UK. When analyzed in combination with spatial data, the spatial patterns of photo density can reveal the preference for different landscape attributes (van Zanten et al., 2016a) or the consequences of land-use change (Sonter, Watson, Wood, & Ricketts, 2016). Currently social media data are also incorporated in the frequently used InVEST ecosystem service models to represent recreation services (Sharp et al., 2016). Yet another step further, Richards and Friess (2015), Tenerelli, Demšar, and Luque (2016), and Oteros-Rozas et al. (2017) not only use the spatial locations of photos but also incorporated the actual content of the photos to make sure only relevant photos are taken into account and to analyze what is actually on the photos to gain more information on landscape preferences. All these studies make stepwise advances in the interpretation of social media for landscape preferences. However, until now the different approaches have not yet been sufficiently combined to provide more in depth understanding of what attracts people in the landscape.

The objective of this paper is to synthesize different approaches to interpret social media photos to pinpoint what it is in a landscape that attracts people. We hypothesize that more insight can be obtained by incorporating both qualitative content of photos as well as the spatial relation to the environment where the photo is taken, to be achieved by combining spatial regression of photo density with systematic content analysis. We used the Kromme Rijn area, a peri-urban agricultural area in the center of the Netherlands as an example case study area.

2. Methodology

The fundamental assumption in this paper is a direct correlation between the density of photos and the aesthetic appreciation of cultural landscapes. A cultural landscape can be described by the combination of its physical components, its management intensity and its cultural value and meaning (Tieskens et al., 2016). These three dimensions determine how a cultural landscape is perceived and valued (Plieninger et al., 2015). We are interested in the causal relation between the physical components of the landscape and its appreciation by people. In prior studies plenty of evidence was found to support the claim that higher densities of Flickr and Panoramio photos suggest higher visitation rates and appreciation of the landscape (Hausmann et al., 2017; Kisilevich, Krstajic, Keim, Andrienko, & Andrienko, 2010; Sun, Fan, Helbich, & Zipf, 2013; Wood et al., 2013). Moreover, multiple studies showed that differences in photo density can partly be attributed to the presence or absence of landscape elements (Gliozzo et al., 2016; van Zanten et al., 2016a). Following van Zanten et al. (2016a) we hypothesize a positive relation between the presence of landscape elements such as water bodies, tree lines and forest, and the aesthetic appreciation of people, measured by photo density.

To test our hypotheses, we downloaded all geo-tagged photos on Flickr and Panoramio in the case study area and performed a negative binomial linear regression to explain photo density with a set of spatial variables consisting of physical landscape attributes, demographics, infrastructure and place specific highlights. Subsequently, photos in areas with large residuals were analyzed using systematic content analysis to derive meaningful inferences about the relation between the landscape and aesthetic appreciation by people.

2.1. Case study area

The Kromme Rijn area (Fig. 1) is a typical peri-urban agricultural area in the center of the Netherlands in the Utrecht province. The area is located on the eastern boundary of Utrecht (city), the fourth biggest city in the Netherlands, making it an ideal location for daily recreation of urbanites. The area contains several towns of which Houten is the largest with just under 50 thousand inhabitants. In total the area was populated by 135.000 people in 2014 (CBS, 2014). Different landscape elements form a variety of typical Dutch landscapes, varying from mosaics with patched forests to wide open pastures on the river bank. Meandering through the case study area is the Kromme Rijn; a stream whose banks were a desirable location for 19th century castles and estates, making the area popular amongst tourists and day-trippers.

2.2. Data

As a proxy for the spatial allocation of aesthetic enjoyment we used the density of unique user uploads of publicly available geo-referenced Panoramio and Flickr content (Tieskens et al., 2016). We chose Panoramio and Flickr because other platforms that have similar potential to contain valuable information have major pitfalls. Twitter users foremost report temporal rather than spatial conditions, while Facebook does not provide the opportunity to download specific geographical content. Instagram does provide an opportunity to download geo-tagged images and is widely used (van Zanten et al., 2016a). However, the Instagram database of geo-tagged photos in Kromme Rijn area only contained 150 geo-tagged photos of the landscape and was, therefore, discarded.

The data were downloaded from the Panoramio and Flickr servers using automated API requests with Python. With these requests we downloaded both the meta-data (geo-tag, upload date and user name) and the actual photo. We downloaded all publicly available data from data sources dated between 2004 and 2017 (Flickr) and 2006 and 2014 (Panoramio). We merged the two data sources as the spatial patterns of Panoramio and Flickr photos are comparable (van Zanten et al., 2016a) and the difference in time period was regarded small in the context of a relatively stable landscape. Next, we used Corine 100-m land cover data for the year 2006 (EEA, 2012) to filter only photos geo-tagged on non-urban land cover (Tieskens et al., 2016; van Zanten et al., 2016a). We chose Corine 2006 as it was closest to the date of the oldest photo in our dataset. Land cover changes between Corine 2006 and Corine 2012 were negligible, making Corine 2006 sufficient for all data. We found no evidence of any significant landscape changes within our case study area that make it necessary to assess the impact of temporal changes in user content across the period. As landscape photos can also be taken from within urban land cover, and to account for inaccuracies in the Corine land cover representation, we included photos from within a 200-m distance of non-urban land cover too. Data were downloaded for the case study area and a buffer of 500 m around the area to account for edge-effects when calculating densities.

The Panoramio dataset contained 7751 photos while the Flickr dataset consisted of 34,401 photos. To avoid bias of very active users we only included one randomly selected photo per user per square kilometer. Both datasets contained a substantial number of photos unrelated to landscape aesthetics. Following Tenerelli et al. (2016) we performed a content analysis of all photos to filter out those photos unrelated to landscape aesthetics. We included only those photos in which the landscape was the main topic. Photos of castles, other buildings, vehicles, or people were excluded. For Panoramio this yielded 3852 photos made by 742 unique users and for Flickr 5579 photos by 898 unique users. On average we manually categorized 1000 photos per hour, totaling to about nine hours work load. We cannot rule out any overlap between the users of the two platforms. Based on the timestamp attached to each photo we determined whether a photo was taken during the winter (December 21–March 20), spring (March 21–June 20), summer (June 21–September 20) or fall (September 21–December 20). In addition to an analysis with all data, we performed a regression analysis for each season separately to capture seasonal differences.

We divided the area in a grid of 100-m cells and calculated the total number of unique user uploads in a 250-m radius neighborhood for each grid cell to account for errors in the location accuracy and the unknown direction of photos. The geotag accuracy error of Flickr rarely exceeds 250 m, while Panoramio geotags are even more accurate (Zielstra & Hochmair, 2013). The final distribution had a mean value of 7.7 unique user uploads per neighborhood of 250 m and standard deviation of 9.6. Accessibility is often mentioned as one of the most important predictors for social media activity. Locations close to roads have often much higher chances of containing social media content than locations not connected by roads (Tenerelli et al., 2016; van Zanten et al., 2016a; Wood et al., 2013). An explorative examination of the dependent variable revealed that 95% of all photos in our database were geo-tagged within 100 m from a road and more than 99% within a distance of 250 m from a road. As our main question is focused on the relation between the aesthetic enjoyment and the landscape rather than the accessibility, we snapped all photos (within a distance of 250 m) to the most recent Open Street Map roads map (OpenStreetMap Contributors., 2017), assuming that photos were taken from a road and included only those grid cells covered by this road network. Photos further than 250 m from a road were discarded. All spatial predictors used in the analysis were also masked to exclude non-road cells.

2.3. Spatial predictors

For each grid cell, we gathered a set of landscape characteristics representing the most important predictors of aesthetic enjoyment mentioned in the literature: hiking and cycling infrastructure, distance to water bodies, recreation sites and landscape elements, and population density (Kienast, Degenhardt, Weilenmann, Wäger, & Buchecker, 2012; Ode & Fry, 2006; van Zanten et al., 2016a; Van Zanten et al., 2014). For population density we used the total population within a certain radius (Kienast et al., 2012). We explored the explanatory power of using a radius of respectively 1 km, 2 km, 5 km 7 km and 10 km. In this test the seven-kilometer radius had the highest explanatory power, and we therefore adopted a seven-kilometer radius. To account for recreation infrastructure, we included the presence of official hiking and cycling routes. We used the location of castles, estates, windmills and churches to account for touristic attraction, based on the listed attractions on the website of the official tourist agency for the Kromme Rijn area (VVV Kromeerijnstreek., 2017; VVV Utrechtse Heuvelrug., 2017). To measure the effect of landscape elements on aesthetic enjoyment we used a map of areas designated with specific nature or landscape conservation management restrictions. We simplified and aggregated the landscape types into six categories: fresh water, grassland, forest, heath and marshland, green linear landscape elements (GLLE) (tree lines, lanes and hedgerows), and orchards (Provincie Utrecht, 2016). For these six categories, as well as for the touristic attraction layer we calculated the inverse of the distance to these features with a maximum of 500 m (i.e. those cells containing the feature have a value of 500, decreasing by 1 every meter away from it to a value of zero in all cells more than 500 m away). The threshold of 500 m delivered the most predictive power in our models, after comparing a threshold of 500, 1000 and 2000 m. To measure the diversity in land cover we used the Shannon Diversity Index. Table 1 shows a list of all predictors used.

2.4. Model estimation

The goal of the model estimation was to predict the density of photos in the best possible way, explaining a maximum fraction of the variance with the least possible number of predictors to estimate the

Table 1
Spatial predictors used for model estimation.

Parameter	Description	Data source
Hiking	Presence of long distance hiking trail	Wandelnet (2017)
Cycling	Presence of national cycling network route	Waypoint Planner (2017)
Monuments	Inverse distance to touristic attractions	Rijksdienst voor het Cultureel Erfgoed (2017), VVV Kromeerijnstreek (2017), VVV Utrechtse Heuvelrug (2017)
Population	Population within radius of 7 km	CBS (2014)
Water	Inverse distance to fresh water	Provincie Utrecht (2016)
Grass	Inverse distance to all grass lands	Provincie Utrecht (2016)
Forest	Inverse distance to all forests	Provincie Utrecht (2016)
Heath	Inverse distance to heath and marsh land	Provincie Utrecht (2016)
GLLE	Inverse distance to green linear landscape elements	Provincie Utrecht (2016)
Orchards	Inverse distance to orchards	Provincie Utrecht (2016)
Shannon	Shannon Diversity Index	
Slope	Maximum slope in cell (in meters)	PDOK (2017)

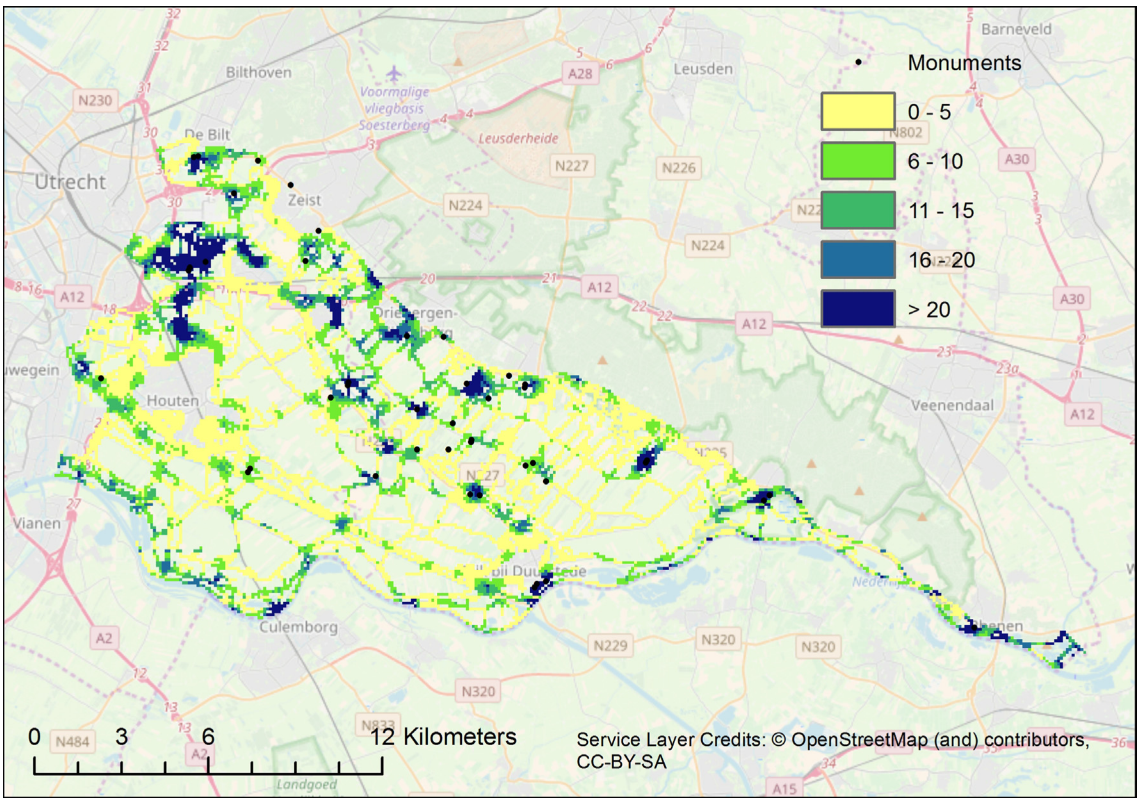

Fig. 2. Observed unique user uploads of landscape photos per hectare in 250-m radius.

influence of landscape attributes on perceived landscape aesthetics controlling for other spatial determinants. We applied generalized linear modeling (GLM) to predict the aesthetic enjoyment in the Kromme Rijn area, as approximated by the unique user uploads of landscape photos (Fig. 2), using the set of predictors as listed in Table 1. To perform the GLM we used the MASS package in R (Ripley et al., 2013). The selection of predictors in the final model was done by both ways stepwise selection of predictors maximizing the residual deviance reduction adjusted for the number of degrees of freedom (adjusted D^2) (Guisan & Zimmermann, 2000). We checked for multi-collinearity by calculating the variance inflation factor (VIF) of each predictor (Guisan & Zimmermann, 2000).

A distribution of non-negative discrete values, such as the photo count in this study, is often fitted with a Poisson distribution (Tenerelli et al., 2016; van Zanten et al., 2016a). However, a Poisson distribution only has one parameter, assuming the mean equals the variance. The variance of the dependent variable ($\sigma^2 = 92.4$) was much larger than the mean ($\mu = 7.7$), indicating severe over-dispersion. Therefore, we used a negative binomial distribution instead, which is very similar to a Poisson distribution but has one extra parameter, allowing the variance to be independent from the mean. To calibrate the relative importance of each predictor we use hierarchal partitioning, using the hier.part package in R (Walsh and MacNally, 2008). With hierarchal partitioning a goodness of fit measure (GOF) is calculated for each possible combination of predictors. The relative importance for each predictor is the average of the difference in GOF of each combination with and without the respective predictor (Chevan & Sutherland, 1991). As our regression is modeled with a negative binomial distribution with a log link we use the log likelihood as GOF (Chevan & Sutherland, 1991).

2.5. Content analysis

To validate the results of the GLM we performed a content analysis of all landscape photos and identified all landscape elements present in each photo. To determine which landscapes elements were on the photos we randomly drew a set of 2000 photos and recoded all visible landscape elements: pasture, arable land, green linear landscape elements (GLLE), forest, monuments (castles, castle gardens, wind mills and towers), water, animals (cattle and wildlife) individual plants, and weather phenomena (significant sky, snow or sunbeams). We compared the relative number of occurrence of each element with the relative regression coefficients.

In some areas the fitted model predicted less unique user uploads than were observed while in other areas the model overestimated the number of unique user uploads. Using the same categories as in the previous content analysis we identified correlations between the presence of landscape elements and the performance of the model by comparing photos in areas with different residuals. We randomly sampled the photos according to the residuals in the grid cell of the photos: 250 photos in areas with residuals over two standard deviations (16 or more unique user uploads more observed than predicted), 250 photos in areas with residuals between one and two standard deviations of residuals, 250 photos with residuals between −1 and 1 standard deviation, and 250 in those areas with more than one standard deviation unique user uploads predicted than observed. In areas with observed values at least one standard deviation under the predicted level we only sampled 136 photos as there were no more photos meeting this criterion.

3. Results

3.1. Model estimation

The observed data show a pattern of several concentrations of photo densities (Fig. 2). The main concentrations, shaded in blue (Fig. 2) were located around known attractions such as Amelisweerd Estate (just east of Utrecht and west of Bunnik), the Amerongen Estate and several other estates and monuments. The areas shaded in yellow and light green (Fig. 2) indicate low densities of landscape social media photos. Stepwise predictor selection resulted in the inclusion of all predictors except for the maximum slope (*Slope*) and the distance to orchards (*Orchard*), which both had an insignificant effect on the dependent variable. There was little to no collinearity among the predictors as the VIF of each predictor was lower than 2 and mutual Pearson's correlations were all under 0.5.

The visual analysis is confirmed by the regression output (Table 2, under "All") as the estimate for the influence of the inverse distance to monuments on the density of photos is relatively high. The value of 0.41 for *Monuments* implies that for each 100 m closer to the *Monuments* features, the natural log of predicted photo density increased by 0.41. The estimates of the first six mentioned variables after the intercept all use inverse distance to features at the same scale and can therefore be

Table 2
Negative Binomial Generalized Linear Model output for all photos and seasons.

Term	Regression estimates									
	Winter		Spring		Summer		Fall		All	
(Intercept)	−1.75	**	−1.77	**	−1.35	**	−2.07	**	−0.26	**
Monuments	0.37	**	0.42	**	0.36	**	0.41	**	0.41	**
Grass	0.11	**	0.1	**	0.07	**	0.06	**	0.07	**
Water	0.15	**	0.09	**	0.14	**	0.12	**	0.14	**
Forest	0.08	**	0.11	**	0.05	**	0.13	**	0.08	**
GLLE	*−0.01*		−0.04		*0.03*	*	0.07	**	0.03	**
Heath	0.04	**	0.08	**	*−0.02*		*0.04*	*	0.06	**
Population	0.14	**	0.16	**	0.15	**	0.16	**	0.14	**
Hiking	0.32	**	0.35	**	0.25	**	0.21	**	0.31	**
Cycling	*0.02*		0.06	**	0.15	**	0.15	**	0.12	**
Shannon	0.33	**	0.38	**	0.09	**	0.41	**	0.24	**
D^2	0.23		0.24		*0.20*		*0.22*		0.31	

** $p < 0.001$.
* $p < 0.01$.

compared with each other. After monuments, the distance to water had the highest impact of these six variables, followed by forest, grass, heath and green linear landscape elements.

The factor predictors cycling and hiking should be interpreted as such that the presence of these elements in a cell increased the log-count of the predicted photo density with the respective value. The presence of hiking trails had a higher explanatory power than cycling paths. All landscape features included in the model have a significant effect on the predicted unique user uploads. The total explained deviance, as compared to the null-model, adjusted for the number of predictors (D^2) was 31% for all seasons.

Table 2 also shows the regression coefficient for the season specific analyses. The most prominent differences between the seasons can be found in the *Forest*, *Hiking* and *Cycling* variables. The distance to forest correlates much stronger with photo density during the spring and fall; hiking trails have a higher photo density during the winter and spring, while cycling paths show the exact opposite pattern.

The hierarchical partitioning outcomes are shown in Fig. 3. The percentage of total explained deviances is projected on the y-axis of Fig. 3 and may reach up to 100. The dark shaded bars (individual) represents the variance explained by the respective variable individually and the light shaded bars (joint) represent the part of the variance explained by each variable but in combination with the other variables. It shows that most explained deviance can be attributed to the inverse distance to monuments. The other main contributors were grasslands, water, forests, hiking trails and the Shannon diversity index. In addition to the deviance explained individually by the predictors a large proportion of the explained deviance can be attributed to the effect of combinations of predictors. Heathland, linear elements, and cycling paths had a smaller contribution to the prediction model.

3.2. Residuals

Visually, the predicted values show a very similar pattern as the observed values with hotspots of predicted unique user uploads at monument locations and along the rivers (Fig. 4). The residuals, however, show a very auto-correlated pattern as is shown in Fig. 5. High concentrations of both high residuals (more observed than predicted) and low residuals (more predicted than observed) are especially located around monuments. This observation is confirmed by the results of the content analysis of residuals.

Fig. 6 shows what percentage of the photos of high negative residuals (− −) to high positive residuals (+ +) contained monuments, water, forests, skies or biodiversity (plants and animals). Monuments, water shows a strikingly similar pattern: in areas where the predicted model had low agreement with the observed values (i.e. in areas with either high positive or high negative residuals) these elements were often present while they were less present when residuals were low. Forest are often present in areas with negative residuals, meaning that the model underestimated the density of photos. The patterns for photos of significant skies or biodiversity are less clear. The other landscape elements (grassland and GLLE) showed no correlation with residuals and are left out of Fig. 6.

3.3. Content analysis

The content analysis of the general sample shows that more than 40% of the landscape photos included water and in 17% of the photos there was forest (Table 3). Despite the high explanatory power of monuments only 11% of the photos showed actual monuments. Grasslands were present in 15% of the photos. We also included weather phenomena, animals, green linear elements (GLLE) and individual plants in our content analysis. Weather phenomena, plants, and animals were present in respectively 12, 8 and 9%. However, we have no spatial data on the presence of these elements.

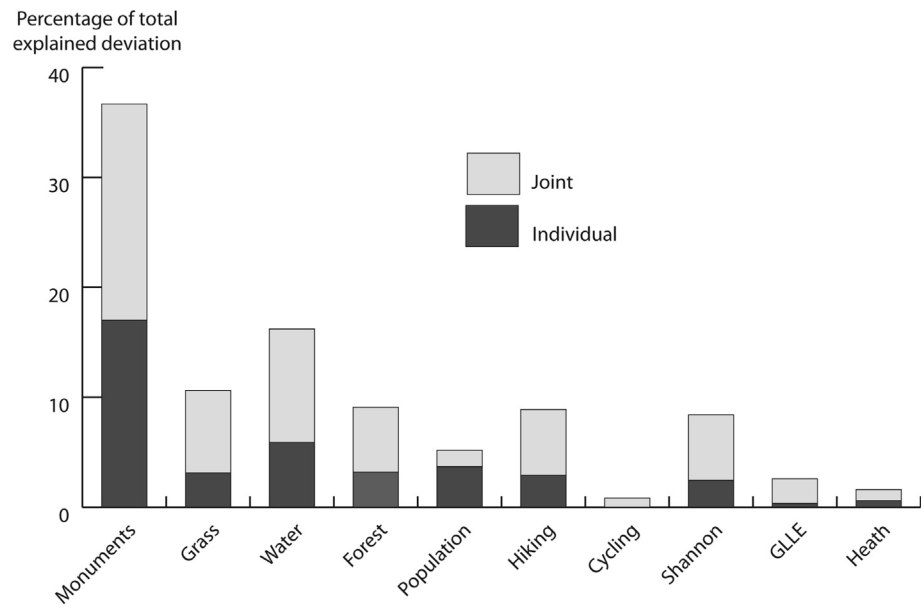

Fig. 3. Relative importance of predictors showing joint effect of explanatory variables with other variables and individual contribution of explanatory variables.

4. Discussion

4.1. Model estimation

The results of this paper indicate that it is indeed possible to go beyond simple counts of social media observations in analyzing landscape preferences. The first promising observation from the model analysis is that our set of predictors increased the predictive power of the model up to almost one third of all deviances as compared to a model with an equal distribution of unique user uploads per grid cell. Also, the visual comparison of the observed values and the predicted values confirm the impression that the set of predictors strongly correlates with the landscape preferences of people in the Kromme Rijn area and overall patterns are well-explained.

Considering the high significance of the landscape elements such as water, grasslands and forest as predictors of landscape preferences we can confidently confirm our hypothesis that landscape preferences are correlated with the presence of specific landscape attributes. One of the interesting results of the analysis is that especially the combination of these elements contributed to the predictive power of the model. This is

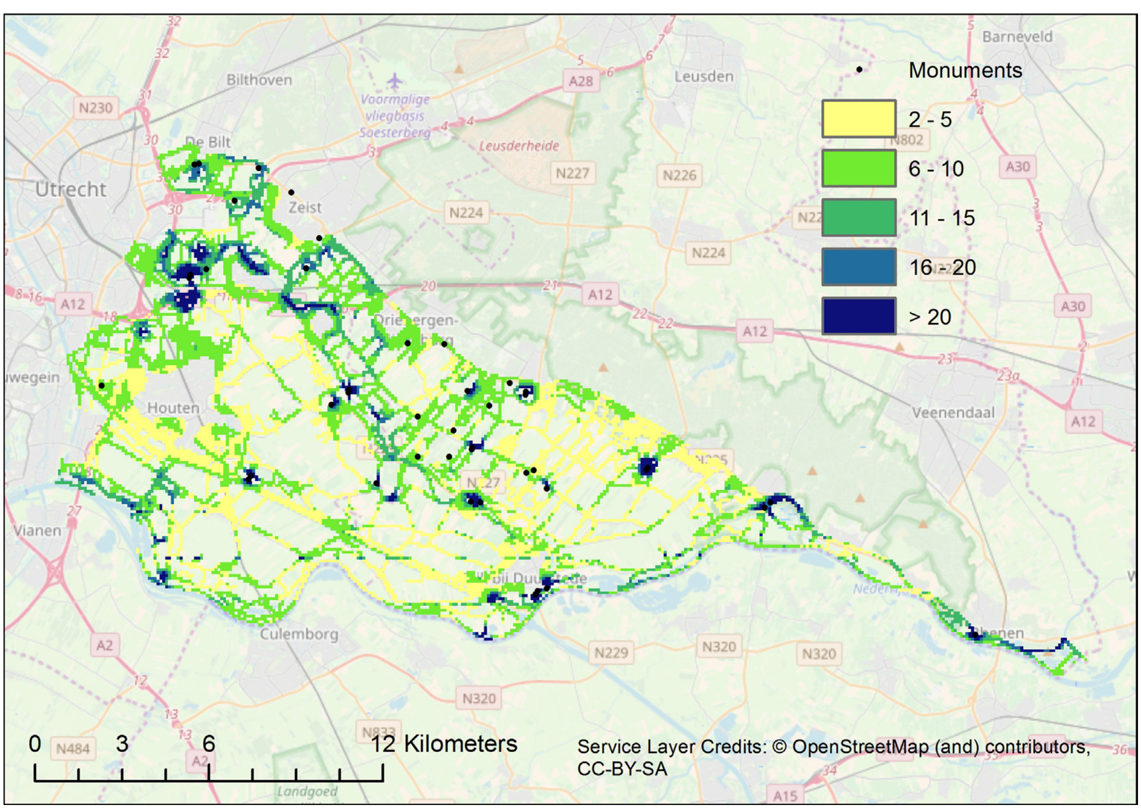

Fig. 4. Predicted values of unique user uploads from model estimation.

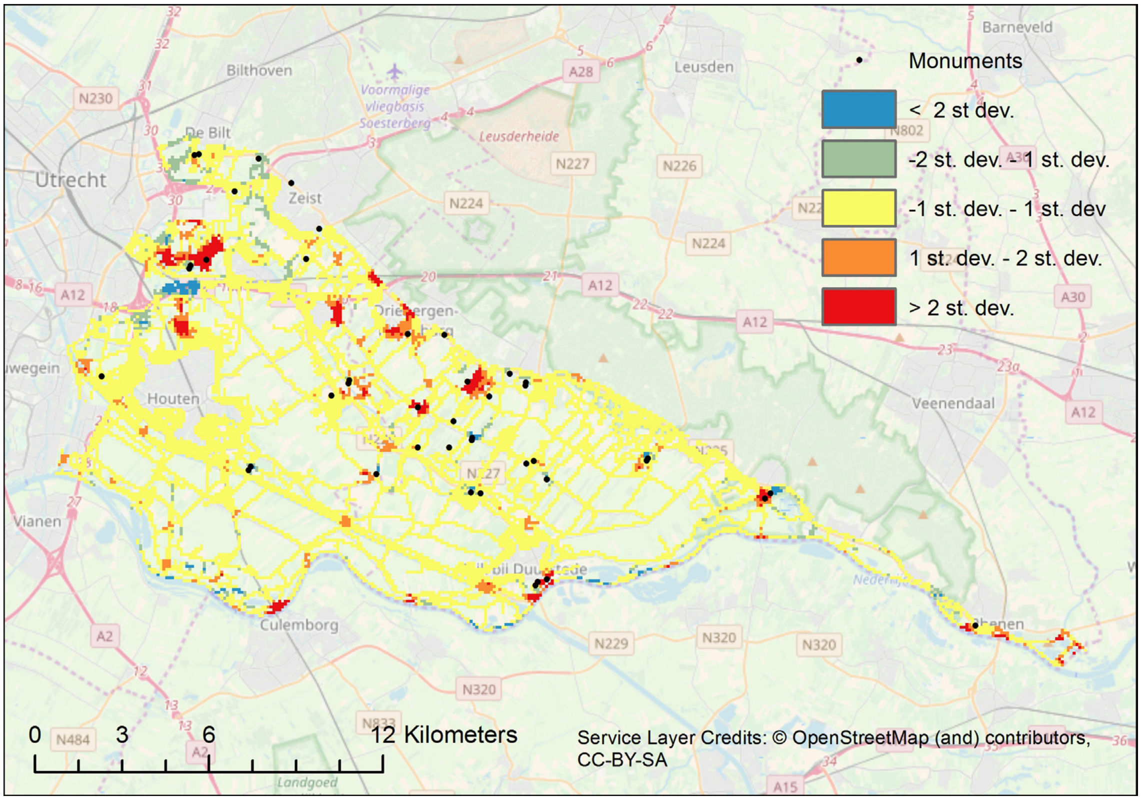

Fig. 5. Residuals of observed values minus predicted values. In the red areas the model predicted too low values while in the blue areas the model predicted too high values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

also confirmed by the positive relation between the Shannon Diversity Index and the density of photos. One of the main determinants of landscape preference is landscape variation; a finding that is confirmed in many stated preference studies (Ode, Fry, Tveit, Messager, & Miller, 2009; Pinto-Correia, Barroso, Surová, & Menezes, 2011; Sayadi, González-Roa, & Calatrava-Requena, 2009), but now also clearly derived from an analysis of social media data.

The results of the GLM and the results of the content analysis show a somewhat different picture. Where monuments have by far the largest influence on the model (Fig. 3), they are only present in 11% of the photos in the content analysis. The presence of water and forest, on the other hand can be interpreted as higher than expected from the GLM outcomes. It shows that the presence of monuments attracts visitors to the landscape but these monuments do not necessarily contribute to the aesthetic value of the landscape themselves. Moreover, the high explanatory power of the monuments variable is striking as photos of

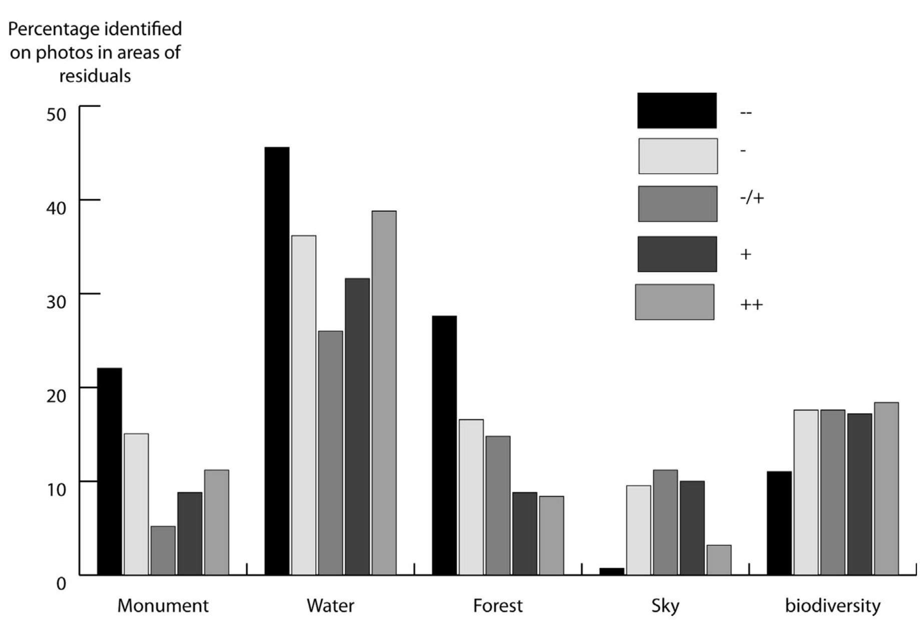

Fig. 6. Percentages of photos containing monuments, water, plants or forests and value within areas with very high positive, positive, close to zero, negative and very high negative residuals.

Table 3
Percentage of photos with presence of each landscape element in random sample (1000) of photos.

	Monument	GLLE	Water	Forest	Weather phenomena	Plant	Animal	Grass
Present in photos	11%	12%	41%	17%	12%	8%	9%	15%

monuments were filtered out. Photographs including monuments were only used in this study if they were part of the landscape as opposed to photos of monuments alone.

The interpretation of the distance to monuments coefficient is therefore complex. If the monument features are interpreted as a landscape element increasing the aesthetic appreciation of the landscape, they can be treated the same way as the more 'natural' features and partly explain landscape appreciation as such. However, some monuments include several hectares of agricultural fields as part of the estate which are subject to specific landscape management regulations (Boon & Schuurman, 2008). It is difficult to determine if the appreciation is related to this typical form of landscape or to the monument itself. In addition, the castles and monuments often serve as museums and therefore attract many visitors that also venture on a walk through the surrounding estate after visiting the museum. Higher photo densities might be explained by higher volumes of people who not necessarily were attracted to the landscape per se but revealed appreciation for the landscape surrounding the main attraction for their visit.

Fig. 6 reveals that monuments are seldom found on photos in areas with low residuals. This indicates that the average coefficient does not accurately reveal the influence of distance to monuments. The density of unique user uploads at some monuments is overestimated while it is the opposite around other monuments. Similar to monuments, some water bodies and forest increase the landscape preference more than others. Just as there is a great variety of different monuments (Wind mills, castles, churches) there are different water bodies such as the river Lek, the Kromme Rijn and several small canals.

The high number of forests on photos in areas where the model overestimated photo density is surprising. An explanation could be that forests are appreciated as elements from a distance as part of a mosaic of land uses or vista, rather than from close by. Similar interpretations have been made by Clay and Daniel (2000) and Hammitt, Patterson, and Noe (1994). Viewshed analysis may be helpful to include in order to include the effect forests have on other landscape elements (Casado-Arzuaga, Onaindia, Madariaga, & Verburg, 2014; Ervin & Steinitz, 2003; Yoshimura & Hiura, 2017).

The seasonality of landscape preferences was surprisingly limited. Apart from the popularity of hiking trails in the winter and spring, as compared to the summer and fall, and the exact opposite for cycling paths, we found no evident difference among seasonal regression estimates. It appears as though the drivers of landscape preference in the Kromme Rijn do no not differ per season.

4.2. Methodological advances

The usage of social media as a proxy for landscape appreciation has three main advantages over traditional stated preference studies. First, they show a trace of actual behavior instead of stated preferences that may suffer from a hypothetical bias. Related to this advantage, second, the photos provided by the users contain information of *in situ* experiences capturing both the local context and experience, as opposed to visualization created by researchers which may give good representations of the biophysical environment but can produce perceptions inconsistent with those in the actual landscape (Daniel & Meitner, 2001). A third advantage of using social media data is that it is relatively cheap and less time consuming than extensive surveys.

Social media have been used as a proxy for landscape appreciation before. However, only very few studies have included content analysis of the actual photos (e.g. Richards and Tunçer, 2018; Richards and Friess, 2015; Oteros-Rozas et al. 2017). The main advantage of content analysis is visual evidence of what it is people appreciate at each location, as compared to a mere location. Additionally, the content analysis provided an extra validation of the correlations found and offered a qualitative insight in concentrations of residuals. A small case study allowed manual content analysis. Advances in machine learning techniques will most likely make it possible to automate the content analysis, making it possible to perform this type of analyses on a larger scale (e.g. Richards and Tunçer, 2018).

However, social media are not the panacea for all landscape preference related research questions as the method has some pitfalls too (Crampton et al., 2013). There is hardly any information available on the users of the social media platforms, which makes inferences more difficult (Tenerelli et al., 2016). Some studies suggest the population of social media users is skewed towards higher educated people (Li, Goodchild, & Xu, 2013). However, methods to identify background information of users and connect them to individual photos are limited. Additionally, social media data often represent a relatively longer timeframe, complicating comparison with static spatial data. Patterns found in "big data" can offer new insights into human behavior but should be treated with caution (Crampton et al., 2013).

The relative simplicity of the method enables research in many case studies and comparisons across landscapes with much less effort; it provides the opportunity for large scale assessment of regional landscape perception. An example of such comparison is Oteros-Rozas et al. (2017). However, while that comparison was only based on simple counts of photo elements and landscape elements, we have shown that formal modeling using spatial predictors can shed additional insights.

The overall method of VGI analysis applied in this paper provides two advances over the cruder photo user day method used in the InVEST model depicting recreation (Sharp et al., 2016). First of all, the filtering of photos before analysis ensures the focus on landscape aesthetic appreciation, while the analysis through generalized linear modeling with content analysis provides insights in the spatial determinants of the observed patterns. An interesting outcome was that the spatial determinants of the photos (i.e. the results of the GLM) do not necessarily correspond to the content of the photos. In other words, where photos are taken does not necessarily equal to what photos are taken. Examples of such differences are provided by the monuments, which were absent from most photos, but were still the most important spatial predictor of photo density; or forests which were often present on photos but its location had only limited influence on the density of photos.

4.3. Application of findings

From an academic perspective this method can be applied to give a detailed summary of the appreciated characteristics of a particular landscape. It provides a quantitative and replicable method to incorporate some aspects of the value and meaning of landscape; a dimension often neglected due to its intangible and hidden nature (Plieninger et al., 2015). Tieskens et al. (2016) provided a very rough characterization method of landscapes on a continental scale using the structure, intensity and value or meaning of the landscape. However, they treated these dimensions as separate pillars in the landscape, while it is the link between those dimensions that characterizes the landscape (Antrop, 2005). In this paper we explicitly made the link between the structure of the landscape (the landscape elements) and the value and meaning (the aesthetic appreciation). The results of the hierarchal partitioning (Fig. 3) reveal which landscape elements contribute to the character of the landscape while the regression coefficients show how strongly each element contributes. One can imagine that two landscapes similar in terms of landscape element configuration but located elsewhere might produce different results. What links the physical structure to the value and meaning is what characterizes the landscape. The link between management intensity and landscape appreciation which is currently still missing from this method, can also be assessed through social media and potentially highlight difference in appreciation of the everyday landscape by visitors and residents (Vouligny, Domon, & Ruiz, 2009).

Additionally, this method also has more direct and applied merits. For instance, the model can be used by local governments assessing how proposed interventions can produce different patterns of landscape appreciation. Also, the construction of recreation infrastructure can be planned more exactly at those locations where landscape appreciation is estimated to be highest. Of course, the same can be done for landscape elements that have an assumed negative effect on the landscape such as wind turbines. Applying the proposed method can for instance reveal where placing wind turbines does the least harm to the predicted aesthetic appeal of the landscape.

5. Conclusion

We proposed a method that can successfully explain spatial patterns of aesthetic appreciation of landscape elements on a local scale. We linked the structural elements of the landscape with its value and meaning, providing an in-depth way of quantifying the appreciation and a characterization of the landscape. However, some elements had a more complex spatial relation to the environment. With the combination of content analysis and GLM we showed how the location and the physical presence of landscape elements both can have their own influence on landscape preferences.

Our results show that social media content provides a powerful indicator of appreciation of the environment and/or visitation rates of places. Additionally, content analysis of the photos can show on a very detailed level what it is that is appreciated in a landscape and can make a more accurate characterization of the appreciated landscape. However, our analysis also suggests that there are more mechanisms that draw people to a place than just its visual attractiveness. The high concentrations of landscape photos around monuments suggest that people often take a photo of a landscape while being there for different reasons, not revealed by social media. Therefore, a broader set of methods remains essential for studying motivations for nature tourism and recreation. Social media data should be incorporated in these methods as it is able to provide evidence of what parts or elements of the landscape are valued, where people are interacting with the landscape, and how these interactions characterize a landscape.

Acknowledgements

This research was supported by the PROVIDE project (Horizon 2020 research and innovation program, grant agreement no. 633838) and the LandSense project, (Horizon 2020 research and innovation program, grant agreement no. 689812), both funded by the European Commission. We thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments.

References

Antrop, M. (2005). Why landscapes of the past are important for the future. *Landscape and Urban Planning, 70*(1–2), 21–34.

Barroso, F. L., Pinto-Correia, T., Ramos, I. L., Surová, D., & Menezes, H. (2012). Dealing with landscape fuzziness in user preference studies: Photo-based questionnaires in the Mediterranean context. *Landscape and Urban Planning, 104*(3–4), 329–342.

Boon, L. D., & Schuurman, S. (2008). De Natuurschoonwet: Borger van Nederlands natuurschoon? *Vakblad Natuur Bos Landschap, 5*(10), 18–19.

Buijs, A. E., Pedroli, B., & Luginbühl, Y. (2006). From hiking through farmland to farming in a leisure landscape: Changing social perceptions of the European landscape. *Landscape Ecology, 21*(3), 375–389.

Casado-Arzuaga, I., Onaindia, M., Madariaga, I., & Verburg, P. H. (2014). Mapping recreation and aesthetic value of ecosystems in the Bilbao Metropolitan Greenbelt (northern Spain) to support landscape planning. *Landscape Ecology, 29*(8), 1393–1405.

Casalegno, S., Inger, R., Desilvey, C., & Gaston, K. J. (2013). Spatial covariance between aesthetic value & other ecosystem services. *PLoS ONE, 8*(6), e68437.

CBS. (2014). Kaart met statistieken per vierkant van 100 bij 100 meter dupdate oktober 2014 (CBS, ed.).

Chevan, A., & Sutherland, M. (1991). Hierarchical partitioning. *The American Statistician, 45*(2), 90–96.

Clay, G. R., & Daniel, T. C. (2000). Scenic landscape assessment: The effects of land management jurisdiction on public perception of scenic beauty. *Landscape and Urban Planning, 49*(1–2), 1–13.

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., et al. (2013). Beyond the geotag: Situating 'big data' and leveraging the potential of the geoweb. *Cartography and Geographic Information Science, 40*(2), 130–139.

Daniel, T. C., & Meitner, M. M. (2001). Representational validity of landscape visualizations: The effects of graphical realism on perceived scenic beauty of forest vistas. *Journal of Environmental Psychology, 21*(1), 61–72.

EEA. (2012). Corine Land Cover 2006 raster data, European Environment Agency (EEA).

Ervin, S., & Steinitz, C. (2003). Landscape visibility computation: Necessary, but not sufficient. *Environment and Planning B: Planning and Design, 30*(5), 757–766.

Gliozzo, G., Pettorelli, N., & Haklay, M. (2016). Using crowdsourced imagery to detect cultural ecosystem services: A case study in South Wales, UK. *Ecology and Society, 21*(3).

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal, 69*(4), 211–221.

Gordon, J. (2013). Comparative geospatial analysis of Twitter sentiment data during the 2008 and 2012 US Presidential elections, University of Oregon.

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling, 135*(2–3), 147–186.

Hammitt, W. E., Patterson, M. E., & Noe, F. P. (1994). Identifying and predicting visual preference of southern Appalachian forest recreation vistas. *Landscape and Urban Planning, 29*(2), 171–183.

Hanley, N., Mourato, S., & Wright, R. E. (2001). Choice modelling approaches: A superior alternative for environmental valuatioin? *Journal of Economic Surveys, 15*(3), 435–462.

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., et al. (2017). Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. *Conservation Letters* n/a-n/a.

Hollenstein, L., & Purves, R. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science, 1*, 21–48.

Kienast, F., Degenhardt, B., Weilenmann, B., Wäger, Y., & Buchecker, M. (2012). GIS-assisted mapping of landscape suitability for nearby recreation. *Landscape and Urban Planning, 105*(4), 385–399.

Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N., Andrienko, G. (2010). Event-Based Analysis of People's Activities and Behavior Using Flickr and Panoramio Geotagged Photo Collections, In 2010 14th International Conference Information Visualisation, pp. 289–296.

Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science, 40*(2), 61–77.

Lothian, A. (1999). Landscape and the philosophy of aesthetics: Is landscape quality inherent in the landscape or in the eye of the beholder? *Landscape and Urban Planning, 44*(4), 177–198.

Ode, Å., & Fry, G. (2006). A model for quantifying and predicting urban pressure on woodland. *Landscape and Urban Planning, 77*(1–2), 17–27.

Ode, Å., Fry, G., Tveit, M. S., Messager, P., & Miller, D. (2009). Indicators of perceived naturalness as drivers of landscape preference. *Journal of Environmental Management, 90*(1), 375–383.

OpenStreetMap Contributors. (2017). Planet dump retrieved from https://planet.osm.org.

Oteros-Rozas, E., Martín-López, B., Fagerholm, N., Bieling, C., & Plieninger, T. (2017). Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. *Ecological Indicators*.

PDOK. (2017). Algemeen hoogte Bestand (AHN-3) 5 meter.

Pinto-Correia, T., Barroso, F., Surová, D., & Menezes, H. (2011). The fuzziness of Montado landscapes: Progress in assessing user preferences through photo-based surveys. *Agroforestry Systems, 82*(2), 209–224.

Plieninger, T., Kizos, T., Bieling, C., Le Dû-Blayo, L., Budniok, M.-A., Bürgi, M., et al. (2015). Exploring ecosystem-change and society through a landscape lens: Recent progress in European landscape research. *Ecology and Society, 20*(2).

Provincie Utrecht. (2016). Natuurbeheerplan 2017 Herziene versie, Utrecht.

Richards, D. R., & Friess, D. A. (2015). A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecological Indicators, 53*, 187–195.

Richards, D. R., & Tunçer, B. (2018). A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecosystem Services*. http://dx.doi.org/10.1016/j.ecoser.2017.09.004 Advanced Online Publication.

Rijksdienst voor het Cultureel Erfgoed. (2017). Rijksmonumenten-kaartlagen.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., Ripley, M. B.

(2013). Package 'MASS', Cran R.
Sayadi, S., González-Roa, M. C., & Calatrava-Requena, J. (2009). Public preferences for landscape features: The case of agricultural landscape in mountainous Mediterranean areas. *Land Use Policy, 26*(2), 334–344.
Schlieder, C., & Matyas, C. (2009). Photographing a city: An analysis of place concepts based on spatial choices. *Spatial Cognition and Computation, 9*(3), 221–228.
Scott, M. J., & Canter, D. V. (1997). Picture or place? A multiple sorting study of landscape. *Journal of Environmental Psychology, 17*(4), 263–281.
Sharp, R., Tallis, H. T., Ricketts, T., Guerry, A. D., Wood, S. A., Chaplin-Kramer, R., et al. (2016). *InVEST User's Guide, The Natural CapitalProject, Stanford University, University of Minnesota.* The Nature Conservancy, and World Wildlife Fund.
Sonter, L. J., Watson, K. B., Wood, S. A., & Ricketts, T. H. (2016). Spatial and temporal dynamics and value of nature-based recreation, estimated via social media. *PLoS ONE, 11*(9), e0162372.
Steen Jacobsen, J. K. (2007). Use of landscape perception methods in tourism studies: A review of photo-based research approaches. *Tourism Geographies, 9*(3), 234–253.
Sun, Y., Fan, H., Helbich, M., & Zipf, A. (2013). Analyzing human activities through volunteered geographic information: Using Flickr to analyze spatial and temporal pattern of tourist accommodation. In J. M. Krisp (Ed.). *Progress in location-based services* (pp. 57–69). Heidelberg: Springer, Berlin Heidelberg, Berlin.
Tenerelli, P., Demšar, U., & Luque, S. (2016). Crowdsourcing indicators for cultural ecosystem services: A geographically weighted approach for mountain landscapes. *Ecological Indicators, 64*, 237–248.
Tieskens, K., Schulp, C. J. E., Levers, C., Kuemmerle, T., Lieskovský, J., Plieninger, T., et al. (2016). Characterization of European cultural landscapes: Accounting for structure, land use intensity and value of rural landscapes. *Land Use Policy*.
van Berkel, D. B., & Verburg, P. H. (2014). Spatial quantification and valuation of cultural ecosystem services in an agricultural landscape. *Ecological Indicators, 37*(Part A), 163–174.
van Zanten, B. T., Van Berkel, D. B., Meentemeyer, R. K., Smith, J. W., Tieskens, K. F., & Verburg, P. H. (2016a). Continental-scale quantification of landscape values using social media data. *Proceedings of the National Academy of Sciences, 113*(46), 12974–12979.
Van Zanten, B. T., Verburg, P. H., Koetse, M. J., & van Beukering, P. J. (2014). Preferences for European agrarian landscapes: A meta-analysis of case studies. *Landscape and Urban Planning, 132*, 89–101.
van Zanten, B. T., Verburg, P. H., Scholte, S. S. K., & Tieskens, K. F. (2016b). Using choice modeling to map aesthetic values at a landscape scale: Lessons from a Dutch case study. *Ecological Economics, 130*, 221–231.
Vouligny, É., Domon, G., & Ruiz, J. (2009). An assessment of ordinary landscapes by an expert and by its residents: Landscape values in areas of intensive agricultural use. *Land Use Policy, 26*(4), 890–900.
VVV Kromeerijnstreek. (2017). Historische plekken.
VVV Utrechtse Heuvelrug. (2017). Historische plekken.
Walsh, C., MacNally, R. (2008). The hier.part Package, in: R package version 1.0.
Wandelnet. (2017). Wandelrouteplanner, HIking trail route planner.
Waypoint Planner. (2017). Fietsroutenetwerk (Fietsersbond).
Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports, 3*, 2976.
Yoshimura, N., & Hiura, T. (2017). Demand and supply of cultural ecosystem services: Use of geotagged photos to map the aesthetic value of landscapes in Hokkaido. *Ecosystem Services, 24*(Supplement C), 68–78.
Zielstra, D., & Hochmair, H. H. (2013). Positional accuracy analysis of Flickr and Panoramio images for selected world regions. *Journal of Spatial Science, 58*(2), 251–273.